

served by various programs, comparison of those persons enrolled in programs with nonparticipants, and assessment of the penetration of programs among high-risk groups. As data and program personnel begin to understand the value of linked files, they can develop records that will facilitate future linkage efforts.

### Conclusion

The linked birth and infant death file, organized as a birth cohort, provides State and national researchers with a unique and powerful tool to analyze infant mortality. The linked file can be readily created to identify maternal and infant factors associated with increased risks of infant mortality. These risk factors can be identified by comparing a State's or an area's infant mortality risks with an appropriate standard. The periodic preparation of a national linked birth and infant death cohort file is an important element in this process (8), because it allows individual areas to make comparisons with national infant mortality risks.

Once the birth outcome file has been created, linkage with other program and service files provides State MCH personnel with an ongoing

system for program evaluation based on firm outcome measures. The creation and routine use of this powerful tool can become a part of the State's process for problem identification, program design, and evaluation.

### References.....

1. Hogue, C. J. R., et al.: Overview of the National Infant Mortality Surveillance (NIMS) project—design, methods, results. *Public Health Rep* 102: 126-138, March-April 1987.
2. Peoples-Sheps, M. D.: Wailing, gnashing of teeth, and trying to evaluate MCH and CC programs. *Network News* 2 (2): 1-4, July 1986.
3. McCarthy, B. J.: The use of birthweight in perinatal surveillance and evaluation. Paper presented at the National Perinatal Care Seminar, Beijing, People's Republic of China, Nov. 1, 1983.
4. Institute of Medicine: Preventing low birthweight. National Academy Press, Washington, DC, 1985.
5. Buehler, J. W., et al.: Birth weight-specific causes of infant mortality, United States, 1980. *Public Health Rep* 102: 162-171, March-April 1987.
6. Centers for Disease Control: National Infant Mortality Surveillance report, 1980. Atlanta, GA. In press.
7. Marks, J., et al.: Variation in State-specific mortality risks. *Public Health Rep* 102: 146-151, March-April 1987.
8. Prager, K., Flinchum, G. A., Johnson, D. P.: Final report of NCHS pilot linked records project. Stage I. *Public Health Rep* 102: 216-223, March-April 1987.

## The NCHS Pilot Project To Link Birth and Infant Death Records: Stage 1

KATE PRAGER, ScD  
GLENN A. FLINCHUM  
DAVID P. JOHNSON

Dr. Prager, a Demographer, and Mr. Johnson, a Survey Statistician, are with the Division of Vital Statistics, National Center for Health Statistics (NCHS). Mr. Flinchum was a Statistician with NCHS before he retired from Federal service.

Tearsheet requests to Dr. Prager, Division of Vital Statistics, National Center for Health Statistics, Rm. 1-44, Center Bldg., 3700 East-West Highway, Hyattsville, MD 20782.

### Synopsis.....

*The National Center for Health Statistics (NCHS) has completed a pilot test of its method to develop national linked files of birth and infant death records. A linked file of the 1982 birth*

*cohort was produced that successfully linked 97 percent of the death records for infants who died in a nine-State area.*

*The method NCHS uses to create national linked files takes full advantage of two existing data sources: the NCHS fully coded natality and mortality files and State files of matched births and infant deaths. For the nine-State pilot area, NCHS obtained computerized linked files from the States and extracted from them the certificate numbers on matching birth and death records. With the use of these numbers, NCHS selected and linked birth and death statistical records from its final natality and mortality files, thus creating new statistical linked records. The initial match rate of 93.2 percent for the project's linked record file was increased to 96.7 percent as a result of efforts by the pilot States to complete the matching of birth and infant death records. Matching in the nine-State linked file appears to be highly accurate, based on the results of two evaluation studies.*

*In the second stage of the project, now underway, NCHS will continue to evaluate and improve State and national linked files for the four birth cohorts of 1983-86. With funding from the Department of Health and Human Services' Office of*

*the Assistant Secretary for Planning and Evaluation, stage 2 will be conducted in collaboration with each of the 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, and New York City, an independent registration area.*

**T**HE NATIONAL CENTER FOR HEALTH STATISTICS' (NCHS) method of developing national linked files of birth and infant death records has been tested and proven to be feasible. A 1982 birth cohort linked file was produced for nine States. Ninety-seven percent of the death records were successfully matched to the corresponding birth records. The test has now been extended to the entire United States and will evaluate national linked files for the birth cohorts of 1983-86. It is being conducted with a view toward incorporating national linked files into NCHS's vital statistics data system.

The pilot test has been carried out with the cooperation of the Association for Vital Records and Health Statistics (AVRHS) and the participating States. Executive staff members of the AVRHS have provided professional support and assistance throughout the development and implementation of the project. Vital statistics personnel from each of the nine pilot States also performed important functions by providing linked data and assisting with the evaluation of the project.

In this paper we present the method NCHS used to compile a linked file for the nine-State area. We also report the completeness and quality of that linked file and their effect on specific rates, and finally, we discuss prospects for NCHS's extending the linked file program nationally.

### **NCHS's Linked Record Project**

The Linked Birth and Infant Death Record Project has been designed by NCHS to produce national, micro-level, linked birth and death data that eventually are to become an integral part of the national vital statistics data system. Although most States currently link birth and death records for infants, national linked data have not been available since NCHS conducted a special study using data for the 1960 birth cohort (1).

Major medical and technological improvements in the care of sick infants that have occurred in the United States since 1960 have contributed to a substantial decline in infant mortality, from 26.0

deaths per 1,000 live births in 1960 (2) to 10.8 deaths per 1,000 live births in 1984 (3). Yet, differentials by race persist. Infant mortality for the black population remains nearly double that for the white population. Furthermore, the rate of improvement in infant mortality has slowed recently, threatening attainment—particularly among blacks—of the national objective of no more than 9 infant deaths per 1,000 live births by 1990 (4). Despite the overall improvement in infant mortality, persistent differentials by race and the recent slowdown in the rate of the decline in mortality highlight the continuing need for data to analyze infant mortality by birth weight and other correlates of premature infant death. Such analyses are needed to guide policy, program planning, and development of interventions to reduce infant mortality by targeting high-risk groups. Linked birth and death records are one obvious source of data for these analyses.

NCHS's linked record project represents Phase II of a two-phased process to develop routine reporting of national, birth-weight-specific infant mortality data. The NCHS project builds on the groundwork laid during Phase I by the National Infant Mortality Surveillance (NIMS) project, which was conducted by the Centers for Disease Control (5). NIMS not only developed interim national linked data but also provided insight into the development of an ongoing system by identifying linkage problems and problems with definitions and explored the usefulness of linked data for program planning and evaluation.

NCHS's pilot project to create national linked birth-infant death files is being conducted in two stages. In stage 1, initiated in 1985, NCHS worked closely with the AVRHS to select the States for the stage 1 pilot test and to develop a methodology. NCHS, AVRHS, and the pilot States jointly evaluated the results of stage 1 in June 1986.

In stage 2, which began in 1986, the pilot test was extended to all 50 States, the District of Columbia, Puerto Rico, Virgin Islands, and New York City, an independent registration area. Four national linked files for the 1983-86 birth cohorts

Table 1. Percent distribution of births and infant deaths and infant mortality rates, by race, in a 9-State area<sup>1</sup> and the United States, 1982

Location	Percent of births			Percent of infant deaths			Infant mortality rates <sup>3</sup>	
	Total <sup>2</sup>	White	Black	Total <sup>2</sup>	White	Black	White	Black
9-State area .....	100.0	84.2	13.9	100.0	72.6	26.0	10.0	21.1
United States .....	100.0	79.9	16.1	100.0	69.9	27.5	10.1	19.6
Difference (absolute).....	...	+ 4.3	- 2.2	...	+ 2.7	- 1.5	- 0.1	+ 1.5

<sup>1</sup> The 9-State area includes Illinois, Indiana, Massachusetts, Michigan, Missouri, New Hampshire, Texas, Vermont, and Wisconsin.

<sup>2</sup> Total includes races other than white and black.

<sup>3</sup> Infant deaths per 1,000 births.

will be produced and their quality and completeness evaluated. Results from these developmental and evaluation activities will provide the guidance and information needed to produce national linked files of births and infant deaths as an integral part of the U.S. national vital statistics system.

The objective of stage 1 was to develop and evaluate a method to produce national linked files and to evaluate their quality and completeness. In stage 1, NCHS used data from the 1982 birth cohort from the following nine States: Illinois, Indiana, Massachusetts, Michigan, Missouri, New Hampshire, Texas, Vermont, and Wisconsin. NCHS selected these States because it believed that the demographic mix would be comparable to that of the United States and that the large number of vital occurrences in the area would provide a reasonable test of a methodology intended for national use. The stage 1 file for these nine States consists of nearly 11,000 infant deaths and more than 946,000 births, or approximately 25 percent of the infant deaths and births that occur annually in the United States.

As shown in table 1, the proportion of vital events for the white population in the pilot area was greater than that in the United States as a whole. Conversely, the proportion of vital events for the black population was lower. For the resident black population in the pilot area, infant mortality in 1982 was higher than it was for the United States: 21.1 deaths per 1,000 live births compared with 19.6. For the resident white population, infant mortality was almost the same as that for the United States (10.0 compared with 10.1).

Within each race group, characteristics of births and infant deaths in the pilot area were similar to those of the United States in terms of such birth characteristics as sex, birth weight, and gestational age; characteristics of the mother; and prenatal care. Thus, when race is controlled, the nine-State

pilot area is reasonably similar to the United States in those characteristics associated with births and infant deaths.

### Methodology

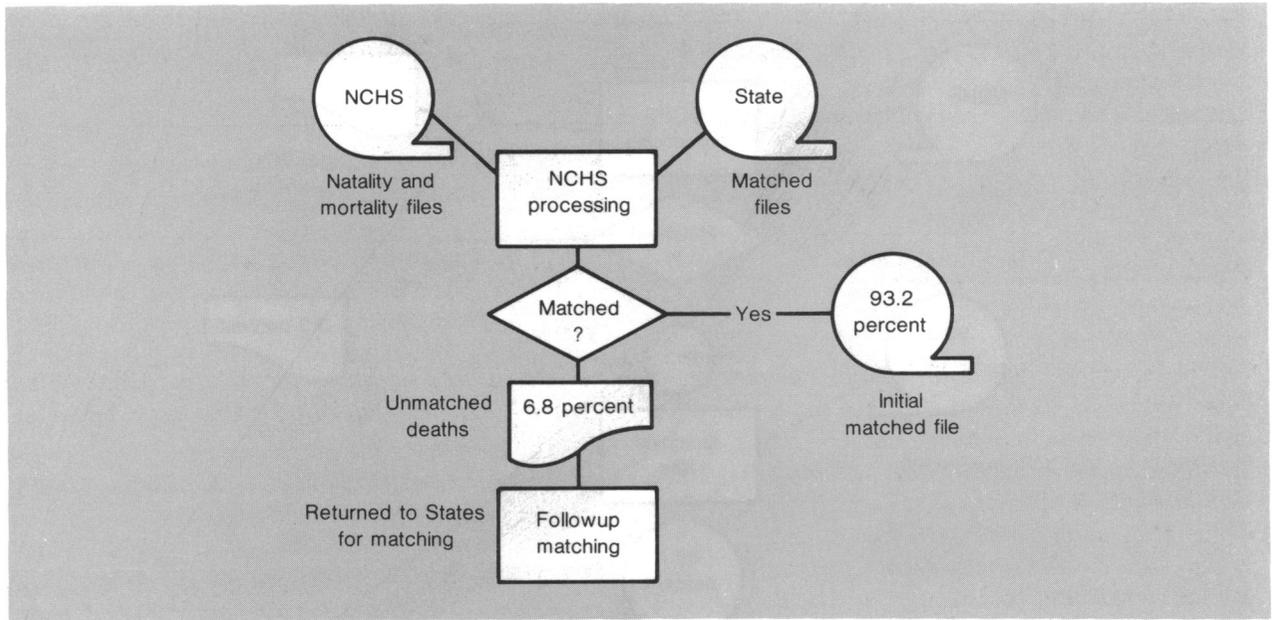
In the methodology being used to create national linked birth and infant death record files, NCHS takes full advantage of two existing data sources: (a) NCHS fully coded natality and mortality files and (b) State matched files. The NCHS files include coded statistical data for births and deaths that have been edited and reviewed and are consistent with information published by NCHS. From these annual mortality statistical files, NCHS identifies infant deaths eligible for the birth cohort linked file.

(Note: An infant death eligible for the stage 1 linked file is the death of any infant born in 1982 in the United States who died in 1982 or 1983 in one of the nine pilot States before his or her first birthday.)

The stage 1 file includes resident and nonresident infant deaths that occurred in the nine pilot States, but the file excludes deaths of infants born outside the United States who died in a pilot State or infants born in pilot States who later died outside the pilot area. The linked file created in the stage 1 pilot study is an occurrence file, rather than a residence file, and it may not be complete for resident births or resident deaths that occurred outside the pilot area. This is an artifact of testing the methodology in a limited number of States, and it will not characterize national files that include infant deaths occurring in all the registration areas in the United States.

The second existing data source is State matched files. Developing a national linked file is facilitated by most States' routinely matching infant deaths to their birth records for legal purposes. When the birth occurs in another State, the interstate certifi-

Figure 1. NCHS methodology to create national linked birth and infant death files: initial processing



cate exchange agreement allows a State to obtain the matching birth certificate for the infant death that occurred in its jurisdiction. Accordingly, the State-of-death and the State-of-birth exchange copies of death and birth certificates to complete these linkages. In addition, if a third State is identified as the State of residence, that State is also sent a copy of the appropriate certificate by the State in which the birth or death occurred.

For the nine-State pilot area, NCHS obtained State linked files on computer tapes and from them extracted the identifying numbers on matching birth and death certificates—but not all the statistical data that were State-coded (fig. 1). For States that did not have matched files on computer tape, matching birth certificate numbers were posted by hand opposite the corresponding death certificate numbers on a list supplied by NCHS. With the use of these identifiers, NCHS selected and linked statistical birth and death records from the NCHS natality and mortality files. Thus, new linked statistical records were created that combine information from birth and death records that previously had been separate. During this initial processing step in stage 1, 93.2 percent of eligible infant deaths were matched.

The remaining 6.8 percent of the eligible infant deaths were not matched, mainly because the State linked files did not include out-of-State births or nonresident deaths. For a few deaths, the State linked file included out-of-State births, but those birth records had been renumbered and, therefore,

could not be identified in the NCHS natality file. For a few other deaths, the matching birth record had been registered after the file was closed and consequently was not available in the NCHS natality file.

To complete the multi-State linked file, NCHS returned copies of infant death records that were not matched during the initial processing step to the nine States for matching. To complete these matches, the pilot State contacted the States of birth identified on the death certificates and obtained matching birth certificates or the original certificate numbers for renumbered birth records. If the matching birth record had been registered late, the pilot States provided NCHS with a copy of the late-registered birth record. In the final step (fig. 2), late-registered birth records were coded and, along with additional interstate linked records, were added to the linked file. As a result of followup matching by pilot States, matching for the stage 1 pilot linked file increased by 3.5 percent, from 93.2 percent to 96.7 percent.

### Characteristics of the Pilot Linked File

The linked file must have a high degree of completeness and quality of matching to be useful to researchers, policymakers, and program monitors. Efforts to increase the completeness of matching have produced measurable results. In this section, we report on the completeness and quality of the matched file and discuss characteristics of

Figure 2. NCHS methodology to create national linked birth and infant death files: final processing

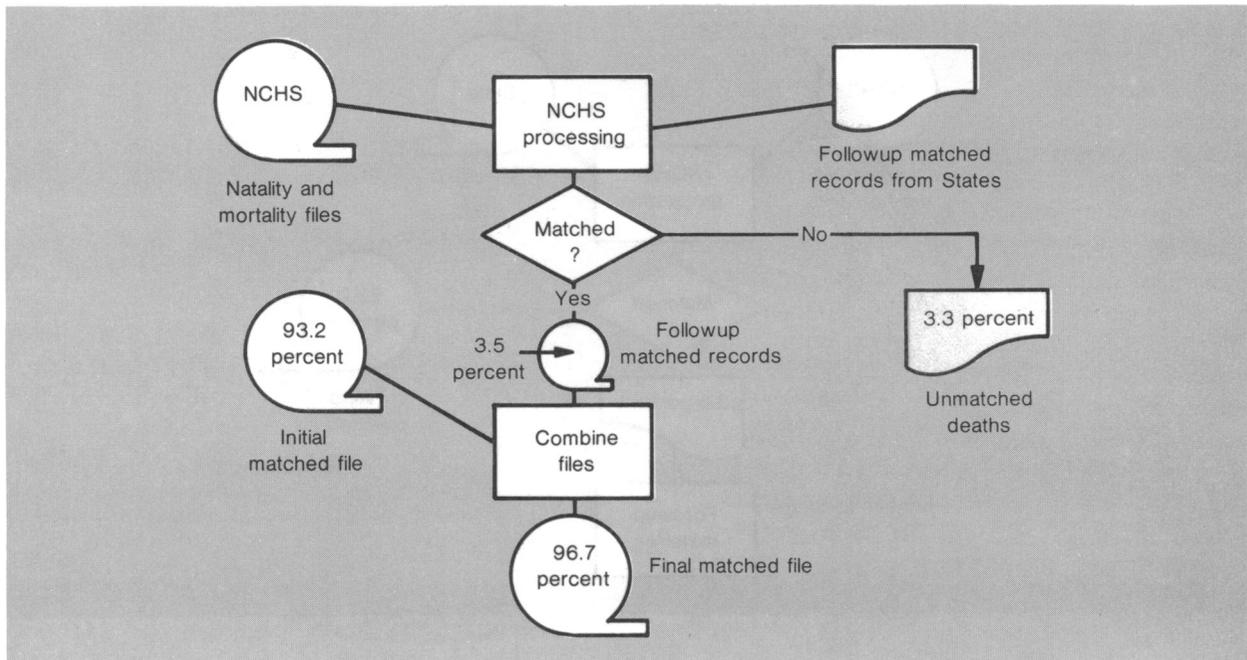
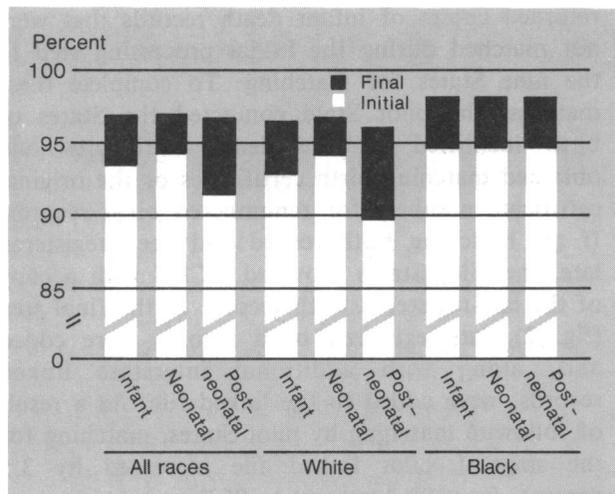


Figure 3. Percent of eligible infant deaths matched by age and race after initial and final processing, 9-State area, 1982 birth cohort



NOTE: An infant death eligible for the stage 1 linked file is the death of any infant born in 1982 in the United States who died in 1982 or 1983 in 1 of the 9 pilot States before his or her first birthday.

considerably after followup matching, generally eliminating the differentials by age and race of infants observed in the initial processing—a fact which underscores the importance of followup.

Figure 3 compares match rates for the total file and for age and race subgroups after initial processing of State linked files and after followup matching. Overall matching increased from 93.2 percent to 96.7 percent. A 1.9 percent initial difference between match rates for black and white infants was reduced slightly after followup. The final match rate was 1.7 percent higher for black infant deaths (97.9 percent) than for white deaths (96.2 percent).

When we considered all races, we initially found a 3.2 percent difference in matching by age at death: 94.2 percent for neonates and 91.0 percent for postneonates. After followup, this difference was virtually eliminated, yielding a final match rate of 96.8 percent for neonates and 96.3 percent for postneonates. Figure 3 also shows that for black infants there was no age differential in matching after either initial or final processing. However, for white infant deaths, initial matching for postneonates was 4.5 percent less (89.6 percent) than for neonates (94.1 percent). Followup reduced this difference to 0.8 percent (95.7 percent for postneonates and 96.5 percent for neonates).

Completeness of matching is strongly associated

nonmatched deaths and the effect of matching completeness on specific rates.

**Completeness of matching.** The final stage 1 linked file is composed of 10,576 linked birth and infant death records, or 96.7 percent of the eligible infant deaths in the 9 States. Match rates improved

Table 2. Infant mortality rates by age and race for all eligible deaths<sup>1</sup> and initial and final files and percent difference between rates for eligible deaths and initial or final files, 9-State area, 1982 birth cohort

Category	Deaths per 100,000 live births		
	Infant (less than 1 year)	Neonatal (less than 28 days)	Postneonatal (28 days-less than 1 year)
<b>Expected infant mortality rates based on all eligible infant deaths:</b>			
All races <sup>2</sup> .....	1,144.3	774.5	369.9
White race.....	991.3	676.2	315.1
Black race.....	2,103.6	1,393.4	710.2
<b>Initial infant mortality rates based on initial processing of State linked files:</b>			
All races <sup>2</sup> .....	1,066.3	729.9	336.4
White race.....	918.5	636.1	282.4
Black race.....	1,990.7	1,318.1	672.6
<b>Percent difference between expected rates and rates based on initial processing of State linked files:<sup>3</sup></b>			
All races <sup>2</sup> .....	6.8	5.8	9.1
White race.....	7.3	5.9	10.4
Black race.....	5.4	5.4	5.3
<b>Final infant mortality rates based on final processing after followup matching:</b>			
All races <sup>2</sup> .....	1,106.3	750.0	356.3
White race.....	953.8	652.3	301.5
Black race.....	2,058.4	1,364.0	694.4
<b>Percent difference between expected rates and rates based on final processing after followup matching:<sup>3</sup></b>			
All races <sup>2</sup> .....	3.3	3.2	3.7
White race.....	3.8	3.5	4.3
Black race.....	2.1	2.1	2.2

<sup>1</sup>Eligible deaths are those of infants born in 1982 in the United States who died in 1982 or 1983 in one of the 9 pilot States before their first birthdays.

<sup>2</sup>All races includes races other than white and black.

<sup>3</sup> $(1 - \frac{R_e}{R_i}) \times 100$ , where  $R_e$  = expected infant mortality rate, and  $R_i$  = initial or final rate.

final rate.

NOTE: The 9-State area includes Illinois, Indiana, Massachusetts, Michigan, Missouri, New Hampshire, Texas, Vermont, and Wisconsin.

with whether the birth and death occurred in the same State or in different States. On the average, for births and infant deaths that occurred in the same State, the match rate was 98.2 percent; for matches in which the birth occurred in another State within the pilot area, the rate was 75.3 percent; for matches in which the birth occurred outside the pilot area, the rate was 53.8 percent. Although the match rate between vital events occurring in two different States is relatively low, fewer than 4 percent of the eligible infant deaths occurred to infants born in another State.

**Unmatched infant deaths.** Followup matching of unmatched infant deaths by pilot States resulted in matching 400 additional records—more than half the initially unmatched death records (6.8 percent unmatched reduced to 3.0 percent). The distribution of characteristics by race and age at death among unmatched deaths differs from the distribution of these characteristics in the matched file. Compared with the matched file, the unmatched

file is composed of proportionately more deaths in the white population (79 percent unmatched compared with 73 percent matched) and fewer deaths in the black population (18 percent unmatched compared with 26 percent matched). In addition, the proportion of postneonatal deaths in the unmatched file (36 percent) is higher than in the matched file (32 percent).

Followup matching reduced the proportion of unmatched death records from 55 to 42 percent for infants whose birth and death occurred in different States. Nevertheless, the unmatched file is still composed of a disproportionately large share of deaths to infants born in another State compared with the matched file. The matched file contains fewer than 4 percent of such deaths.

A total of 364 records remained unmatched after final processing for the following reasons: exclusion of deaths of nonresident infants from State matched files; failure of interstate certificate exchange; inability to obtain original certificate numbers for renumbered records; and underregistration

*'Efforts expended in matching additional unmatched records not included in the initial State linked files paid off in terms of boosting the matching completeness by 3.5 percent. Analysis of these results also shows that there is still room for improvement.'*

of births. A few infant death records were ineligible for the 1982 birth cohort linked file because they represented deaths of infants born outside the United States or contained misreported information.

**Effect of matching completeness on specific rates.** Incomplete matching causes underestimation and possibly a bias on specific infant mortality rates. Efforts to complete the matching of the file are important for achieving infant mortality rates free of biases and thus useful for analysis and monitoring. This section shows the impact of followup matching on infant mortality rates.

Infant mortality rates by age and race were computed for the initial and final files and compared with rates that would be expected if all eligible infant deaths were included (table 2). The initial file is based on State linked files provided to NCHS; the final file includes additional followup matched records.

Before followup matching, differences between matched and expected rates were larger for the white than the black population and for postneonatal than neonatal deaths. Final processing reduced these differences considerably as shown in table 2. For white postneonatal deaths, the initial 10.4 percent difference between matched and expected rates was reduced to 4.3 percent; for white neonatal rates, the 5.9 percent difference was reduced to 3.5 percent.

In general, matching was more complete for the black than the white population. In both the initial and final files, the differences between expected and matched file rates for the black population were smaller than for the white population. Also, for the black population, there was very little difference between expected and matched file rates for neonates and postneonates.

**Quality of matching.** Quality of matching was evaluated in two groups of records: a randomly selected quality control sample of records from each State linked file and a select group of records where the possibility of poorer quality matching was expected to be greater, that is, among records containing statistical inconsistencies. On the basis of results from both of these evaluations, matching in the nine-State linked file appears to be highly accurate.

The purpose of the evaluation was to estimate the true match rate for the State matched files and to determine whether statistically undetectable mismatched records will be a problem in a national linked file.

Matching accuracy was evaluated in a quality control sample of 100 records randomly selected from State linked files numbering over 400 records. A larger ratio of records was selected from smaller State linked files. The birth and death records composing the linked records were copied from microfilm and examined for accuracy. The results of this manual record review indicated that matching in these State matched files is highly accurate. A confidence interval of two standard errors below the observed rate yields a match rate of no lower than 99.6 percent. Thus, one may conclude that the true match rate is better than 99 percent.

In the second part of the evaluation, records with inconsistencies in race, sex, date of birth, and State of birth between the birth and death portions of the matched record were rejected, copied from microfilm, and examined. We found 551 records (5.4 percent of the initial matched file) with at least one inconsistency. Of this group, only 17 records were incorrectly matched, yielding a mismatch rate of 3.1 percent among linked records with inconsistent statistical data. Thus, even among this small group of records at highest risk of including mismatched records, the match rate was still 96.9 percent.

## Discussion

States use matching procedures that vary according to frequency of matching, method of matching (manual or computer), and comprehensiveness of the resulting file. Because the quality and completeness of the NCHS national linked file are dependent on State matching practices, it is useful to identify some practices that produce better results. From written documentation of State matching practices, several commonalities have emerged:

1. Matching weekly or monthly facilitates feedback between the matching program and registration of vital events. Annual matching does not appear to be integrated well with vital registration.

2. When records are matched at least monthly, the unmatched deaths suggest unregistered births or out-of-State births that need followup.

3. Manual matching appears to produce better results, perhaps because of the greater matching frequency used by the States that match manually. Seven of the nine pilot States matched records manually, whether the files numbered under 100 or more than 2,000 annual deaths. Two States matched records annually by computer and, in general, did not have as complete a matched file.

4. Verification of matching by any of a variety of techniques and adjudication of discrepancies is important to ensure the validity of the match itself and to ensure the consistency of the information on the two records.

5. More completely matched files result when they are comprehensive for all occurrences, regardless of residence. Several States explicitly exclude State occurrences to out-of-State residents, thus producing a residence-only State linked file. From a national perspective where control and processing are based on State of death, residence files are a limitation, resulting in incomplete matching that requires followup efforts to complete the matching.

6. Finally, the practice of some States of renumbering out-of-State certificates reduces the apparent completeness of a national linked file because only the original number is on records in the NCHS statistical files. This problem could easily be corrected if each State were to include the original number in the file along with the new number assigned to the out-of-State record.

**Continuation of the pilot linked record project at NCHS, stage 2.** Results from stage 1 of the pilot project demonstrated the feasibility of creating a multi-State linked file using this methodology. Efforts expended in matching additional unmatched records not included in the initial State linked files paid off in terms of boosting the matching completeness by 3.5 percent. Analysis of these results also shows that there is still room for improvement. While a 97 percent match rate is a good start, matching completeness should be significantly improved in the stage 2 national pilot project.

There are at least two reasons to expect improvements in matching completeness in stage 2.

First, we hope to identify State practices and procedures that produce better results and offer this knowledge and experience to other States to improve their own practices. Second, stage 2 will be conducted on a national basis. NCHS will encourage all States to complete their linked files and to cooperate with other States in completing files through interstate exchange of records.

With funding from the Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services, NCHS and the States will continue to evaluate and support improvements in State and national linked files for the four birth cohorts of 1983-86. Stage 2 will be carried out with the collaboration of each of the 50 States, the District of Columbia, Puerto Rico, Virgin Islands, and New York City, an independent registration area.

Strong matching programs can contribute significantly to improved data quality through better registration, more complete interstate exchange of certificates, and more pertinent information on infant deaths. Improved match rates over the span of this pilot project will enhance the value of the linked data base for research, policy, and program purposes.

#### References.....

1. Armstrong, R. J.: A study of infant mortality from linked records by birth weight, period of gestation, and other variables, United States. *Vital Health Stat* [20] No. 12. DHEW Publication No. (PHS) 79-1055. National Center for Health Statistics, Hyattsville, MD, 1972.
2. National Center for Health Statistics: *Vital statistics of the United States, 1960. Vol. II, Mortality, Part A.* U.S. Government Printing Office, Washington, DC, 1963.
3. National Center for Health Statistics: *Advance report of final mortality statistics, 1984. Monthly Vital Statistics Report 35, No. 6, supplement (2).* DHHS Publication No. (PHS) 86-1120. Hyattsville, MD, Sept. 26, 1986.
4. Public Health Service: *Healthy people: the Surgeon General's report on health promotion and disease prevention, 1979.* DHEW Publication No. 79-55071. U.S. Government Printing Office, Washington, DC, 1979.
5. Hogue, C. J. R., Buehler, J. W., Strauss, L. T., and Smith, J. C.: *Overview of the National Infant Mortality Surveillance (NIMS) project—design, methods, results.* *Public Health Rep* 102: 126-138, March-April 1987.